

Mozgást jelentő igék argumentumszerkezetének korpuszalapú vizsgálata

Lázár Bernadett¹, Szabó Martina Katalin^{1,2,3}, Vincze Veronika^{4,5}

¹PrecognoX Informatikai Kft.

²Szegedi Tudományegyetem, Szláv Intézet, Orosz Filológiai Tanszék

³MTA TK „Lendület” RECENS Kutatócsoport

⁴MTA-SZTE Mesterséges Intelligencia Kutatócsoport

⁵Szegedi Tudományegyetem, Általános Nyelvészeti Tanszék

Kivonat: Dolgozatunkban az *átjön/bejön*, valamint az *átmegy/bemegy* igekötős igepárok argumentumszerkezeit vizsgáljuk a Szeged és a Dívány korpusz adatai alapján. A kutatás célja az, hogy számítógépes eszközökkel, korpuszadatok alapján feltárjuk a kiválasztott igék argumentumszerkezeteinek kvantitatív és kvalitatív sajátosságait, majd vizsgálataink eredményeit összevessük különböző szótárak szócikkeivel és a szakirodalom vonatkozó megállapításaival. A vizsgálattal tehát egyrészt az igei sajátságok megismerésére törekszünk, másrészt szeretnénk megtudni, hogy a szótárak, valamint a szakirodalmi adatok a valós nyelvhasználatot tükrözik-e. Mindezzel összefüggésben a dolgozat fontos célja az is, hogy a lexikológiai kutatásokhoz adalékul szolgáljon. A munka során azt is meg szeretnénk vizsgálni, hogy mutatkozik-e eltérés a két, szöveganyagukat tekintve jelentősen különböző korpusz adataiban.

1 Bevezetés

Dolgozatunkban az *átjön/bejön*, valamint az *átmegy/bemegy* igekötős igepárok argumentumszerkezeit vizsgáljuk a Szeged Korpusz [16] és a Dívány korpusz [10] adatai alapján. A dolgozatot a következőképpen építettük fel:

A vizsgálat első lépéseként két szerző vonatkozó eredményeit [13, 14, 5] tekintjük át, majd azokat összevetjük a magyar értelmező szótárak adataival. Az elméleti áttekintés után számítógépes nyelvészeti eszközökkel, korpuszadatok alapján vizsgáljuk meg az igéket. Ehhez először röviden bemutatjuk a felhasznált korpuszokat, majd a lekérdezéssel kapott eredmények alapvető statisztikai adatait. Azt is megvizsgáljuk, hogy mutatkozik-e valamiféle eltérés a két, szöveganyagukat tekintve jelentősen különböző korpusz adataiban. Végül a kapott eredményeket összevetjük a szótárak adataival és az elméleti irodalom tárgyalt megállapításaival.

2 Szakirodalmi áttekintés

Az igekötős igék vizsgálatával számos tanulmány foglalkozik, az elméleti és az alkalmazott nyelvészet oldalán egyaránt.

Az egyik elméleti nyelvészeti megközelítési mód az emberközpontúság fogalma felől magyarázza az igekötős igéket. Ez azt jelenti, hogy tárgyakat, élettelen dolgokat (az elmélet terminológiája szerint az ún. *nem viselkedőket*) *viselkedőkre* jellemző tulajdonságokkal ruházunk fel (vö. [5]). E megközelítés alapján a világfelfogásunkat alapvetően a mozgás határozza meg, amit a nyelvhasználat is tükröz. Radden [7] szerint egy entitást úgy érzékelünk, mint amely mozgásban van akkor, amikor lokalitása érzékelhető, látható változáson megy keresztül. A mozgásnak van egy kezdőpontja és egy végpontja, ahová a kezdőponttól egy úton keresztül jutunk el, így adott az irányultság is (vö. [7]).

A fenti elmélettel összhangban van Lakoff és Johnson [6] mozgásértelmezése, amelyben a következő elemekkel operálnak a szerzők: *forrás, ösvény, cél* és *irányultság*. Úgy vélik, hogy “bármikor elmozdulunk valamerre, van egy hely, ahová érkezünk, egy összefüggő pontok sorozata, mely összeköti a kezdeti pontot a végponttal, valamint létezik irányultság” [6].

Szilágyi N. Sándor [11] ugyancsak hasonló megállapításokat tesz a magyar nyelv vonatkozásában. Véleménye szerint “a magyarban az igekötőnek talán a legfontosabb szemantikai szerepe éppen az, hogy megjelenítse, explicitté tegye a nyelvi világ térstruktúrájában való elmozdulásokat. Alkalmassint ezzel függ össze az is, hogy az igéket befejezett aspektusúvá teszi” [11].

A térbeli elmozdulás alapján az igekötők abban is különböznek egymástól, hogy a KEZDET–ÚT–VÉG sémából mit ragadnak meg. A *be-* és a *ki-* igekötők például a kezdetre vagy a végpontra vonatkoznak, a *fel-* és a *le-* igekötők az irányultságra, az *át-* igekötő utalhat a kezdetre és a végre (1a) vagy az út egy kiemelt helyére, ami által a mozgásban levő valamilyen relációba kerül (1b).

- (1) a. Átment Európából Amerikába.
b. Átszökött a kerítésen.

A továbbiakban azokat a megállapításokat tekintjük röviden át, amelyek konkrétan az általunk vizsgált igekötős igékre vonatkoznak. A vizsgálat során elsősorban Szili Katalin [13, 14], valamint Imre Attila [5] dolgozataira támaszkodtunk, majd a megállapításokat összevetettük az értelmező szótárak adataival.

Szili Katalin [13, 14] azt vizsgálta, hogy milyen jelentésmódosulásokon keresztül válik az irányjelentéssel bíró *be-* igekötő (pl. *bemegy, bejön, befut*) tiszta perfektiváló elemmé (pl. *bevall, bebizonyít*). Elemzésében egyszerre használ szintaktikai és szemantikai vizsgálati szempontokat, és a különböző fogalmi típusokra tartályokként utal. Vizsgálatai kiindulópontját olyan tárgyatlan és tárgyas igékkel álló szintagmák képezik, amelyekben az igekötő klasszikus irányjelentésben szerepel. Ezeknek az igekötős igéknek a sajátossága, hogy az ige környezetében megjelennek a *be* vonzatának tekintett *hova?* kérdésre válaszoló *-ba/-be* ragos helyhatározók, hozzájárulva az igekötő irányjelentéséhez. Mellettük még olyan argumentumok is

megjelenhetnek, mint például a *-hoz, -hez, -höz* határozóragos nyelvi elemek vagy a *honnan?* kérdésre felelő, ablatívuszi esetragos bővítmények.

A *be-* igekötős igék kapcsán Szili [13, 14] az argumentumok minősége alapján, a belső tereknek négy nagy típusát különíti el: helyiség jellegűek (pl. ház, lakás, szoba), konténerszerű tárgyak (pl. doboz, táska, zseb), élő és élettelen testek (pl. ember, fa, kenyér), valamint csoportok, halmazok. Szili [13, 14] alapján az általunk vizsgált *bemegy* és *bejön* mozgásigék leginkább az első csoport elemeivel állnak. Tekintsük az alábbi példákat! (A példákban a vizsgált elemeket félkövér szedéssel emelem ki.)

- (1) a. **bemegy** a házba
b. **bejött** a szobába

Ezeket a szerző [13, 14] a helyiség jellegű helyekkel álló igék csoportjának nevezi és megállapítja, hogy környezetükben a már említett *honnan?* és *hova?* kérdésekre felelő határozók fordulnak elő. Az így létrejövő vonzatstruktúrában megmarad a *be-* igekötő az eredeti irányjelentésében, és az alanyi argumentum bekerül valamilyen helyiségszerű térbe.

A különböző nyelvekben ugyanaz a tér máshogy jelenhet meg. Az olyanokat, mint például a *tér* a magyarban sík, lapszerű területként fogjuk fel, míg a parkot, erdőt vagy az utcát, amelyben házak állnak, zárt belső területként.

A „HIVATAL, INTÉZMÉNY, KÖZPONT” típust a szerző átmeneti csoportként definiálja. Átmeneti, hiszen képzeletünkben a hivatalok, intézmények és maguk az azoknak helyet adó épületek (a konkrét zárt helyek) összemosódnak. Ezeket vagy konkrét intézményként fogjuk fel, vagy csak az azoknak helyet adó épületekként gondolunk rájuk, például

- (2) **Bejön** ma a főnök?

A Magyar nyelv értelmező szótára [1] ezt a példát a következőképpen magyarázza: “a beszélő megérkezik hivatalába, belső helyiségben lévő munkahelyére, ahol a beszélő is tartózkodik”. Szili [13, 14] megközelítése tehát egybevág az értelmező szótár definíciójával: mind a két megközelítés kulcsfontosságúként emeli ki azt a tényt, hogy BENT helyről beszélünk.

Szintén zárt térről raktározzuk tudatunkban a minket körülvevő dolgokat, tulajdonainkat (vö. [13, 14]), például

- (3) Csinos kis összeg **jött be** a gyümölcsből.

Ugyancsak zárt térről jelennek meg az emberek kisebb-nagyobb időleges csoportosulásai is, például

- (4) A legtöbb tag 6-8 holddal **jött be** a szövetkezetbe.

Az *átjön/átmegy* igekötős igék vizsgálata során Imre Attila [5] eredményeire támaszkodtunk. Imre [5] alapján ahhoz, hogy „ÁT” helyzetéről beszélhessünk, észlelési feltételekre van szükség. Észlelési feltételnek a következőket tekinti:

- a) Adott a kezdet–út–vég” séma és a mozgás;
- b) Feltételezünk egy azonos térben elhelyezkedő, ún. stabilit és egy ún. mobilt, amelyek egymáshoz viszonyított helye változik a mozgás során;
- c) A viszonyítás alapja a stabil, ehhez képest a mobil közeledik felé, kapcsolatba kerül vele, majd megszűnik a viszony. Fontos, hogy a mobil a stabil által az úton kijelölt helynek egyik oldaláról a másikra kerül.

Imre [5] a következő stabil-típusokat különbözteti meg: anyagszerű stabilok, más jellegű stabilok (a rész-egész viszonyt kialakító), valamint az emberi test, mint stabil (lehet BENNE hely is, RAJTA hely is). Tekintsük Imre Attila az *átmegy/átjön* igékötös igékre vonatkozó megállapításait e rendszer alapján!

Az igékötös igék hordozhatnak konkrét, valamint metaforikus értelmet, és ezt gyakran a stabil természete határozza meg. Például, ha a mobil a stabil határvonalai között van, és az egyik vonaltól eljut a másikig, akkor konkrét értelmű kifejezésről beszélünk:

(5) **Átment** a szobán.

A példában a mobil a stabil határai között mozog.

A határátlépés egy speciális esetének tekinthetjük a színek közötti átmeneteket. Tekintsük az alábbi példát!

(6) **Átmegy** vörösből rózsaszínbe.

A példában jól érzékelhetően elkülönül a két végpont egymástól, azonban magát az átmenetet, tehát azt, hogy pontosan mi van a két végpont között, jóval kevésbé lehet meghatározni.

Lehetséges olyan eset is, amikor a mobil a stabil egy úgynevezett virtuális határvonalán jut át úgy, hogy nem érintkezik, például

(7) Most **megy át** a város fölött.

Dichotómiás szerkezetekben (pl. kicsi-nagy, egyik-másik stb.) két stabilról beszélünk, amelyekkel a mobil érintkezik, például

(8) **Átment** a kicsi szobából a nagyba.

Mint ahogy korábban említettük (l. fentebb), Imre [5] szerint a stabilok nem csak anyagszerűek, hanem viselkedők is lehetnek. Ez azt jelenti, hogy egyes stabilok emberi tulajdonsággal felruháztak is lehetnek. Ilyen viselkedő stabilit látunk a következő példában:

(9) A bárányhimlő **átment** a testvéreire.

3 Az argumentumszerkezetek problémaköre a számítógépes nyelvészetben

Az igei argumentumszerkezetek vizsgálata mind elméleti, mind számítógépes nyelvészeti szempontból fontos probléma. A nyelvtechnológia oldalán azért is különösen jelentős, mert az igekötős igék argumentumszerkezetének automatikus elemzése még nem teljes mértékben megoldott, bizonyos részproblémáira megoldás még nem született. (A jelen dolgozatnak nem célja, hogy ezekkel a problémákkal foglalkozzon, részletesebben l. [7]).

Az alábbiakban csupán megemlítünk néhányat a magyar nyelvre vonatkozó nyelvtechnológiai kutatások és fejlesztések közül.

A vonzatkeretek egy fontos vizsgálati eszköze a Magyar Nemzeti Szövegtárban lévő, online elérhető kutatóeszköz, melynek segítségével az igei vonzatkereteket, az igei bővítményszerkezeteket tudjuk kvantitatívan tanulmányozni (vö. [8]). A vonzatkeretek elsajátításának modellezését és az igei vonzatkeretek gépi tanulásának más nyelvekre jól működő megközelítéseit, statisztikai módszereit alkalmazva (vö. [2]), a gépi tanulási görbéket a gyereknyelvi adatokon számított tanulási görbékkel vetették össze. A kutatás fő irányvonala az argumentumstruktúrák elsajátításának számítógépes modellezése volt.

A vonzatkeretek számának, a magyar igék hosszának és gyakoriságának összefüggését vizsgálták a Szeged Dependency Treebank (vö. [16]) rövid üzleti híreket tartalmazó alkorpuszának összegyűjtött adatai alapján (vö. [17]). Feltételezték, hogy a leggyakoribb igék rendelkeznek a legtöbb vonzatkeretekkel, a legrövidebb igék a leggyakoribbak és a legrövidebb igéknek vannak a legtöbb vonzatkeretei. A vizsgálatot kiterjesztették a teljeskörű vonzatkeretekre is, majd ezeket összehasonlították egy elméleti alapon épített vonzatkeret-lexikonban található vonzatkeretekkel.

4 A vizsgálati korpuszok bemutatása

A lekérdezésekhez a Dívány [10] és a Szeged korpuszt [16] alkalmaztuk. A vizsgálatunkhoz olyan szöveganyagokat kerestünk, amelyek szignifikánsan eltérnek egymástól, és az egyiket referenciának tekinthetjük. Ebből a megfontolásból tehát a sokféle szövegtípust tartalmazó Szeged Korpuszt [16], valamint a téma- és műfajspecifikus, termékvélemény-szövegekből álló Dívány korpuszt [10] választottuk.

A Dívány korpusz egy magyar nyelvű szentimentkorpusz, amelyet termékvélemény-szövegekből hoztunk létre kutatási és fejlesztési céllal [9]. A korpusz szöveganyagát a [<http://divany.hu/>] honlap termékvéleményeiből állítottuk össze. A korpusz fragmentum- és aspektusszintű annotációval rendelkezik [10]. Az adatbázis összesen 154 vélemény-szöveget, 17 059 mondatot és 251 202 tokent (központozással) tartalmaz.

A Szeged Korpusz egy kézzel annotált, 1,2 millió szavas magyar nyelvi korpusz. A korpusz létrehozásának egyik fő szempontja volt, hogy az a lehető legkülönbözőbb műfajú és témájú szövegeket tartalmazza, így jött létre az a hat különböző témakörből

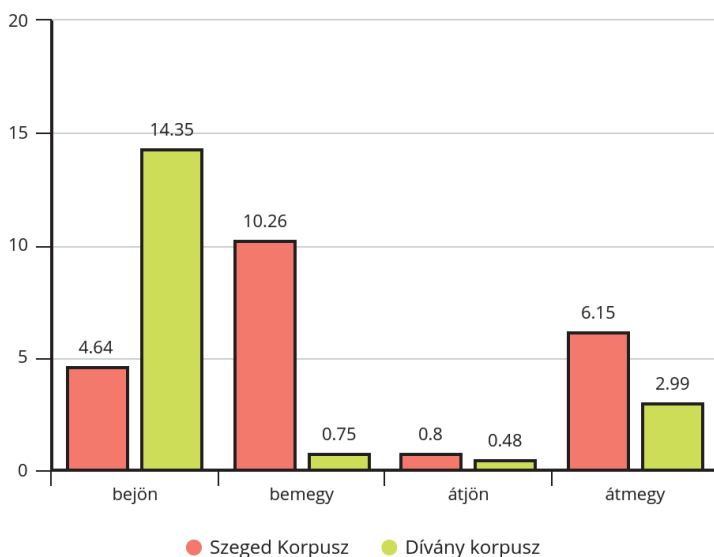
álló adatbázis, témakörönként körülbelül 200 ezer szó terjedelemben. A Szeged Korpusz szépirodalmi szövegeket, 14–16 éves tanulók fogalmazásait, újságcikkeket, számítástechnikai témájú és szövegeket, valamint gazdasági és pénzügyi rövidhíreket tartalmaz.

5 A korpuszadatok elemzése

A Dívány és a Szeged Korpuszból a *jön/megy* igéket, valamint azok igekötős változatait kérdeztük le. Az eredmények az egyes igealakokat, valamint az azokhoz tartozó argumentumokat tartalmazták, az MSD-kódrendszer szerinti jelölésekkel. Az adatok „full valency” [14] értelemben, azaz az összes bővítményükkel szerepelnek. A „full valency” azt jelenti, hogy a kötelező és szabad bővítményeket egyaránt az igei vonzatkeret részének tekintjük [3].

Az eredmények feldolgozásánál külön-külön elemeztük a *jön/megy* igék adatait igekötő nélkül, valamint az *átjön-bejön* és *átmegy-bemegy* igéket is.

A lekérdezések eredményeit az alábbi diagram szemlélteti:



Ami a Dívány korpusz adatait illeti, az összesen 251 202 szövegszóban 343 alkalommal fordult elő a *jön* és a *megy* ige. Kézzel megvizsgáltuk a kapott adatokat, és azt tapasztaltuk, hogy a igeik leggyakrabban átvitt értelemben szerepelnek a korpuszban, például

- (10) Várom, hogy kiérzzek gyümölcsízeket, de nem **megy**.
- (11) A második vonalban **jön** az enyhe, értelmezhetetlen mellékíz.

Amennyiben a *bejön* és az *átjön* igeik előfordulásait a *bemegy* és az *átmegy* előfordulásaival vetjük össze, azt látjuk, hogy az előbbiek előfordulása jóval gyakoribb az utóbbi ige pár előfordulásánál (megközelítőleg hatszor gyakrabban fordul elő). Ha az ige kötők szerint vetjük össze az ige párok előfordulásait, érdekes eredményt kapunk: amíg a *bejön* jóval gyakoribb a *bemegy* igenél, addig az *át*-esetében éppen az ellenkezőjét figyelhetjük meg, tehát az *átmegy* lesz a gyakoribb a *bemegy* igenél.

A lekérdezés egy meglepő tapasztalata az volt, hogy az MNSZ2 adatai gyakorinak várt *bemegy* ige mindössze 1 adattal szerepelt a korpuszban. Nem meglepő módon azonban az *átjön* ige, amely az MNSZ2 korpuszban szintén csekély számmal fordult elő, a Dívány korpuszban is mindössze 1 adattal szerepelt:

- (12) Olyan mintha tök náthás lennék, és esélytelen, hogy ízhez jussak, aztán végül mégis **átjön** valami.
- (13) Ha **bemegyek** a boltba, biztos nem ezt vásárolok meg.

A Szeged Korpusz 1,2 millió szavas adatbázisa a *jön* és a *megy* igeiket összesen 1599 alkalommal tartalmazta. Ha a két korpusz lekérdezési eredményeit összevetjük egymással, azt látjuk, hogy a vizsgált igeik találati száma összességében arányos a korpuszok méretével.

A Szeged Korpuszban a vizsgált ige kötös igeik egymáshoz képesti gyakorisági adatai nem egyeznek a Dívány adataival. A Szeged Korpuszban ugyanis az *átmegy* és *bemegy* igeik fordulnak elő nagyobb gyakorisággal az *átjön* és a *bejön* igeiknél.

A két korpusz adatai közötti talán legszembeötlőbb eltérés az, hogy amíg a Dívány korpuszban a *bejön* ige a *jön* összes előfordulásának majdnem a 15%-át teszi ki, addig a Szeged Korpuszban ugyanez mindössze 4,64%. Ezt az eltérést véleményünk szerint akár a következő sajtóság is okozhatja: mivel a Dívány korpuszt termékvélemények alkotják, a *bejön* ige átvitt értelemben, pozitív értékelés kifejezésére is gyakorta használhatják a szövegekben (pl. *nekem nagyon bejön ez a telefon!*). Feltevésünk igazolására megnéztük a Dívány korpuszban a *bejön* ige összes előfordulását, és azt tapasztaltuk, hogy a 30-ból 26 esetben használták az ige átvitt értelemben, tehát pozitív vélemény kifejezésére, például

- (14) Savanykás, ez nekem jobban **bejön**, mint az eddigiek
- (15) Kicsit édesebb, mint ami nekem **bejön**

Fontos megemlíteni, hogy a szótárak szócikkeiben nem szerepel a *bejön* ige esetében ez a jelentés, ugyanakkor azt látjuk, hogy a tényleges nyelvhasználatban

gyakori. A szótárak tehát ebben az esetben nem korrelálnak a valós adatokkal, ami azok kiegészítését, revízióját sürgeti.

Ezzel ellentétet látunk akkor, ha a *bemegy* ige adatait vizsgáljuk. Ez ugyanis a Szeged Korpuszban a *megy* összes előfordulásának több mint a 10%-át teszi ki, míg a Díványban alig fordul elő.

Az *átjön* mindkét korpuszban nagyon ritka, legfeljebb csupán 1%-át teszi ki az *jön* ige összes előfordulásának. Az *átmegy* ige ugyanakkor kétszer olyan gyakran fordul elő a Szeged Korpuszban a Dívány adataihoz képest.

A továbbiakban azt vizsgáljuk, hogy a lekérdezett igék hány argumentummal fordulnak elő az egyes korpuszokban. Azt is megfigyeljük, hogy a *jön* és a *megy* igék argumentumszámán változtat-e, ha igekötővel állnak, és ha igen, hogyan módosítja azt.

A Dívány korpuszban a *jön* ige esetében a leggyakoribb az első három pozíció kitöltése, amely az összes előfordulás 58%-át teszi ki. A *jön* ige legmagasabb argumentumszáma 8, amely csupán egyetlen esetben fordul elő, és az ige *be*- igekötős alakjának az adata:

(16) Így pedig nagyobb az esély arra, hogy hat, de nekem hosszú távon sajnós nem **jött be**, de ez egyéni dolog.

A 8 bővítményt tehát három kötőszó, egy szuperesszívuszi esetű főnév, egy határozói bővítmény, egy tagadószó, egy kérdőszó, valamint egy alanyi bővítmény adják.

A *megy* ige leggyakrabban 3 argumentummal fordul elő (44%). Az ige legmagasabb argumentumszáma 7, amely a *jön* ige esetéhez hasonlóan mindössze egyetlen egyszer fordul elő, és a *rá*- igekötős ige adataként szerepel. A 7 bővítmény ebben az esetben egy kötőszó, három határozói bővítmény, egy adesszívuszi és egy szublatívuszi esetű főnév, és egy időhatározói bővítmény adják.

Amennyiben az igekötős igék argumentumszámait vizsgáljuk, a következőket tapasztaljuk: A Dívány korpuszban a *bejön* ige legmagasabb argumentumszáma 6, azonban mind az 5, mind a 6 argumentum nagyon ritka, az összes előfordulásnak csupán a 10%-át tesz ki. Ugyancsak ritka a 3 és a 4 argumentummal való előfordulás. 26 esetben két argumentumú az ige, ami az összes eset (30) 86%-a. Ezt tekinthetjük tehát a leggyakoribb esetnek a Dívány korpusz alapján. Az *átjön* ige legmagasabb argumentumszáma 3, amelyek az első két kitöltöttségi helyen határozói, a harmadik helyen pedig alanyi bővítmények. Azt mondhatjuk tehát, hogy a *jön* ige esetében az igekötő megléte jelentősen redukálja az argumentumok mennyiségét.

A *bemegy* ige esetében ugyanazt tapasztaltuk, mint az *átjön* igénél. Legmagasabb argumentumszáma 3, amelyek rendre illatívuszi esetű főnévi, valamint határozói szófajú bővítmények. Az *átmegy* ige legmagasabb argumentumszáma 5, ez azonban csupán 2 esetben fordul elő. A *megy* ige esetében ugyanakkor az 5 argumentum kimondottan gyakori.

Ugyanezek a vizsgálatok a Szeged Korpuszon az alábbi eredményeket hozták: A *jön* és a *megy* igék legmagasabb argumentumszáma a 7 volt. Az alapigék esetében az argumentumszámok a Szeged Korpuszban nem mutatnak jelentős eltérést a Dívány korpuszban mért eredményeinkhez képest.

A *bejön* ige legmagasabb argumentumszáma 6, amely egyetlen egyszer szerepel a korpuszban. A Szeged Korpusz *bejön* igéjének adatai az argumentumszámokat illetően hasonlóak a Díványéhoz, itt is a két argumentum a leggyakoribb, amely a 79%-át teszi ki az összes előfordulásnak. Az *átjön* ige legmagasabb argumentumszáma 3, amely ugyancsak egyszer szerepel a korpuszban. Ezek az eredmények szintén egyezést mutatnak a Dívány korpuszban kapott eredményeinkkel.

Az *átmegy* és a *bemegy* igék esetében a legmagasabb argumentumszám 5. A *bemegy* iginél szintén a 2 argumentum a leggyakoribb, amely 89%-át teszi ki az összes előfordulásnak.

Ha összevetjük a *jön* és *megy* igéket azok igekötős alakjaival, azt tapasztaljuk, hogy ezek argumentumai a *bejön* ige kivételével jelentősen leredukálódtak.

Összevetve a két korpusz fentebbi adatait megállapítható, hogy mindkét korpuszban az igekötő megléte, illetve annak hiánya jelentősen változtatott az argumentumok számán. Mind a két korpuszban a *jön* és *megy* igék magasabb argumentumszámmal fordultak elő igekötős változataiknál, és az igekötős alakok átlagos legmagasabb argumentumszáma (a *bejön* ige kivételével) 3 volt.

A kutatás eredményei azt mutatják, hogy a vizsgálati korpuszok szövegműfaji sajátosságai jelentősen nem befolyásolták az argumentumok számát. A vizsgált igék argumentumszámai ugyanis a két korpuszban nem mutattak szignifikáns eltérést, 4 esetben teljesen meg is egyeztek az adatok.

Azt is megnéztük, hogy az általunk választott igekötős ige párok mely vonzatokkal állnak a leggyakrabban. Azokat az igéket, amelyekre valamely korpuszban nagyon kevés az adat, itt nem elemezzük.

A Dívány korpuszban a *bejön* ige esetében azt tapasztaltuk, hogy a 30 előfordulásból 18 esetben kapcsolódott az igehez datívuszi esetű argumentum (17a), míg a Szeged Korpuszban az összes 29 előfordulásból csupán 11 alkalommal (17b).

- (17) a. Akinek ez **bejön**, azzal nagy gond van.
b. Szlovákiának **bejött** az euró.

Úgy véljük, ez a sajátosság is összhangban áll korábbi megállapításunkkal, amely szerint a Dívány korpusz termékvélemény-szövegeiben a *bejön* igtét átvitt értelemben, a tetszés kifejezésére használják a leggyakrabban.

A Szeged Korpuszban az *átmegy* ige 11 alkalommal állt illatípuszi raggal, ami az összes előfordulás (60) 18,33%-át adja, például

- (18) Péter **átment** a szomszédba.

Ugyancsak 11 alkalommal társult hozzá allatípuszi rag, ami összhangban van a szakirodalom és a szótárak megállapításaival, miszerint az ÁT-igekötő jelölhet RAJTA és BENNE helyet is, például

- (19) Anna **átment** Julcsihoz.

A Szeged Korpuszban a *bemegy* ige 39 alkalommal állt illatípussszal, ami 39%-át teszi ki az összes előfordulásnak, például

(20) Ha senki nem ér rá, ki **megy be** a boltba?

Elmondható tehát, hogy vizsgálati eredményeink ebben a vonatkozásban összhangban állnak a szakirodalmak és a szótárak definícióival. Egyedüli különbséget a *bejön* és ige pozitív értékelő funkcióban való használatát jelenti, amelyre sem a szakirodalmakban, sem a szótárak definícióiban nem találtunk példát.

6 Összegzés

Dolgozatunkban az *átjön/bejön*, valamint az *átmegy/bemegy* igekötős igepárok argumentumszerkezeteit vizsgáltuk a Szeged [16] és a Dívány korpusz [10] adatai alapján.

A vizsgálat első lépéseként a hazai elméleti nyelvészeti irodalom néhány fontosabb eredményét (vö. [13, 14, 5]) tekintettük át. Az elméleti áttekintés után lekérdeztük a vizsgált igekötős igéket a Dívány és a Szeged Korpuszból, és az adatok alapján feltártuk az igék nyelvhasználati sajátosságait a valós szövegekben. Az elemzés keretében mind az előfordulási gyakoriságokat, mind a vonzatokra vonatkozó sajátosságokat megvizsgáltuk, valamint összevetettük egymással a két korpusz adatait. Mindemellett a nyelvhasználati adatokat összevetettük a szakirodalmi eredményekkel is.

A vizsgálatok alapján a következő legfontosabb megállapításokat tettük: A Dívány korpuszban a *bejön* jóval gyakrabban fordul elő annál, mint ahányszor a másik vizsgálati korpuszban szerepel, emellett jelentősen nagyobb a frekvenciája az összes többi igekötős igénél is. Úgy véltük, hogy ez a Dívány korpusz sajátosságát mutatja. A *bejön* igét ugyanis a termékvélemény-szövegekben valószínűleg gyakran használják átvitt értelemben, tetszés kifejezésére. Az adatok kézi ellenőrzése igazolta a hipotézisünket, ugyanis feltárta, hogy az összes előfordulásból 26 valóban ebben az értelemben szerepelt. Ez az eredmény különösen fontos, hiszen sem a szakirodalom, sem a szótárak szócikkei nem említik az igének ezt a jelentését.

A Szeged Korpusz kvantitatív adatai nem álltak összhangban a Dívány fentebbi adataival. A *bejön* és az *átjön* igékkel szemben ugyanis a *bemegy* és az *átmegy* igék előfordulása volt a gyakoribb. Az MNSZ2 [8] gyakorisági adatai alapján csekély számúnak várt *átjön* ugyanakkor mindkét korpuszban nagyon ritka, legfeljebb csupán 1%-át teszi ki a *jön* ige összes előfordulásának.

Megvizsgáltuk az argumentumok számát is az igekötős és igekötő nélküli igék esetében. Azt tapasztaltuk, hogy a *jön* és a *megy* igék alapvetően több argumentumokkal állnak igekötő nélkül, az igekötők ugyanis mind a két korpuszban és rendre minden ige esetében leredukálták az argumentumok számát. A két korpusz eltérő szöveganyaga azonban vizsgálataink alapján nem befolyásolta az argumentumok számát az egyes igék esetében. A *megy*, *bejön*, *átjön* és *átmegy* igék esetében a legmagasabb argumentumszámok a két korpuszban megegyeztek.

Azt is megnéztük, hogy az általunk választott igekötős igepárok mely vonzatokkal állnak a leggyakrabban. Azokat az igéket, amelyekre valamely korpuszban nagyon

kevés az adat, nem elemeztük. A *bejön* ige kapcsán a datívuszi esetű argumentum volt szembevetendő, amely az összes előfordulás 60%-ában volt megtalálható, míg a Szeged Korpuszban ugyanez csupán 37%-ot tett ki. Gyakori volt azonban az illatívuszi és allatívuszi esetű argumentum is, amely az *-át* igeekötő esetében alátámasztotta a szakirodalmak meghatározásait.

A kutatás további tervezett lépései a következők: Egyrészt a jelen vizsgálatot szeretnénk más ige párokra, és lehetőség szerint más korpuszokra is kiterjeszteni. Másrészt, hogy hasonló, elméleti szempontú és korpuszalapú vizsgálatokat szeretnénk végezni az orosz nyelvre vonatkozóan is, majd a kapott eredményeket összevessük a magyar nyelvre vonatkozó eredményeinkkel.

Köszönetnyilvánítás

A jelen kutatást a PrecognoX Informatikai Kft., az Emberi Erőforrások Minisztériuma Új Nemzeti Kiválóság Programja, az MTA Társadalomtudományi Kutatóközpont „Lendület” RECENS Kutatócsoportja, az EFOP-3.6.1- 16-2016-00008 azonosítójú, EU társfinanszírozású projekt, valamint az European Research Council (ERC) támogatta az Európai Unió „Horizon 2020” nevű, 648693-as számú kutatási és fejlesztési programja keretében. Vincze Veronika kutatásait az Emberi Erőforrások Minisztériuma ÚNKP-17-4 kódszámú Új Nemzeti Kiválóság Programja támogatta.

Bibliográfia

1. A magyar nyelv értelmező szótára. I. Akadémiai Kiadó, Budapest. (1978)
2. Babarczy A., Serény A., Simon E.: Magyar igei vonzatkeretek gépi tanulása. In: Tanács A., Szauder D., Vincze V. (szerk.) VI. Magyar Számítógépes Nyelvészeti Konferencia, Szegedi Tudományegyetem, Szeged. (2009) 333-342.
3. Čech, R., Pajas, P., & Mačutek, J.: Full valency. Verb valency without distinguishing complements and adjuncts. *Journal of Quantitative Linguistics*, 17(4), (2010) 291–302.
4. É.Kiss K.: Egy igeekötőelmélet vázlata. *Magyar Nyelv*, 100. (2004) 15-43.
5. Imre A.: Az ÁT viszony jelentésszerkezete a magyar nyelvben. Egyetemi szakdolgozat. Kézirat. Kolozsvár. (1999)
6. Lakoff, G., Johnson, M.: *Metaphors We Live By*. Chicago, London, The University of Chicago Press. (1980)
7. Lázár B., Szabó M.K.: Mozgást jelentő ige argumentumszerkezetének korpuszalapú vizsgálata, XII. Alkalmazott Nyelvészeti Doktoranduszkonferencia, (2018) absztrakt benyújtva
8. Oravecz Cs., Váradi T., Sass B.: The Hungarian Gigaword Corpus. In: *Proceedings of LREC 2014*, 2014.
9. Radden, G.: Motion Metaphorized. The Case of Coming and Going. In: *Studies in Linguistics III, A Supplement to the Hungarian Journal of English and American Studies*, Ed.by Béla Korponay and Péter Pelyvás, Debrecen. (1994)
10. Sass B.: Vonzatkeretek a Magyar Nemzeti Szövegtárban. In: Alexin Z., Csendes D. (szerk.): IV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2005). Szegedi Tudományegyetem, Szeged. (2005) 257-264

11. Szabó M.K., Vincze V.: Egy magyar nyelvű szentimentkorporusz létrehozásának tapasztalatai. In: Tanács A., Varga V., Vincze V.: (szerk.) XI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2015). Szegedi Tudományegyetem, Szeged. (2015) 219-226
12. Szabó M.K., Vincze V., Hangya V.: Aspektusszintű annotáció és szentimentet módosító elemek egy magyar nyelvű szentimentkorporuszban. In: Tanács A., Varga V., Vincze V. szerk. XII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2016). Szegedi Tudományegyetem, Szeged. (2016) 174–182.
13. Szabó M.K., Vincze V., Simkó K., Varga V., Hangya V.: A Hungarian Sentiment Corpus Manually Annotated at Aspect Level. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). European Language Resources Association (ELRA), Portoroz, Szlovénia. (2016). 2873-2878.
14. Szilágyi N.S.: Hogyan teremtsünk világot?, Erdélyi Tankönyvtanács, Kolozsvár (1996)
15. Szili K.: A be igekötő jelentésváltozásai I. Magyar Nyelvőr 129. (2005) 151-164.
16. Szili K.: A be igekötő jelentésváltozásai II. Magyar Nyelvőr 129. (2005) 282-299.
17. Tolcsvai N.G.: Az igekötő + ige szerkezet szemantikája. Nyelvtudományi Közlemények 109. (2013) 187-226.
18. Vincze V., Szauter D., Almási A., Móra Gy., Alexin Z., Csirik J.: *Hungarian Dependency Treebank*. In: Proceedings of the Seventh Conference on International Language Resources and Evaluation. European Language Resources Association, Valletta, Málta (2010) 1855-1862.
19. Vincze V.: Valency Frames in a Hungarian Corpus. Journal of Quantitative Linguistics, 21:2. (2014) 153-176.
20. Forrás: <http://www.divany.hu>